iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82127

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

AI-Powered DPR Quality Assessment and Risk Prediction System

Medishetty Swapna[1], Kasarla Narasimha Reddy[2], Lagudu Ramkishore[3], Panthini Sathvik[4]
[1]*Assistant Professor, Department of Information Technology, MVSR Engineering College*
[2, 3, 4]*Department of Information Technology, MVSR Engineering College*

Abstract: ***The evaluation of Detailed Project Reports (DPRs) in government infrastructure projects is traditionally a manual and time-consuming process, often leading to inconsistencies, delays, and cost overruns. This paper proposes an AI-Powered DPR Quality Assessment and Risk Prediction System that automates the analysis of DPR documents using Natural Language Processing (NLP) and Machine Learning techniques. The system extracts key information such as project objectives, budget, and timelines, evaluates document quality based on predefined criteria, and predicts risks such as cost overruns and delays using models like XGBoost and LightGBM. Additionally, Explainable AI techniques and Monte Carlo simulations are used to provide transparent insights and probabilistic risk analysis. The system improves efficiency, ensures standardized evaluation, and supports data-driven decision-making in government project management.***
Keywords: DPR Analysis, Machine Learning, NLP, Risk Prediction, Explainable AI, Government Projects

I. INTRODUCTION

The increasing complexity of infrastructure projects in India has made the evaluation of Detailed Project Reports (DPRs) a critical task. Traditional DPR evaluation methods rely heavily on manual analysis, which is time-consuming, error-prone, and lacks consistency. Government projects often face delays and cost overruns due to poor-quality DPRs and lack of early risk identification.
The proposed system provides an AI-based solution that automates DPR evaluation using NLP and machine learning techniques. It analyzes document structure, extracts key information, evaluates quality, and predicts potential risks. This approach improves efficiency, reduces manual effort, and enables faster and more reliable decision-making for government authorities.

II. LITERATURE REVIEW

Existing systems in project evaluation focus either on document analysis or risk prediction. Some systems use NLP techniques to extract information from documents, while others use machine learning models to predict project risks based on historical data. However, most existing solutions lack integration between document understanding and predictive analytics.
Recent research highlights the importance of combining AI techniques for better decision-making. Systems using machine learning algorithms such as Random Forest and XGBoost have shown promising results in risk prediction. Similarly, NLP-based systems have improved document analysis. However, there is a need for a unified system that integrates both approaches.

A. Gap in Research Findings

From the literature, it is evident that existing approaches are largely fragmented and domain-limited. NLP-based systems focus primarily on information extraction, while machine learning models rely on structured datasets for prediction, creating a disconnect between document understanding and risk forecasting. Additionally, most systems do not incorporate region-specific regulatory compliance, which is critical in the Indian context, especially for ministries like MDoNER. The absence of probabilistic risk modeling and explainable AI mechanisms further limits their applicability in government decision-making processes. The proposed system addresses these gaps by integrating document intelligence, quality scoring, compliance validation, machine learning-based risk prediction, and explainable AI into a unified framework, enabling a comprehensive and transparent DPR evaluation process.

III. EXPERIMENTAL METHOD / PROCEDURE / DESIGN

The development of the AI-Powered DPR Quality Assessment and Risk Prediction System is carried out through a structured process involving document processing, natural language processing, machine learning, and web application development. The system is implemented as a web-based platform that allows users to upload DPR documents and receive automated analysis and risk predictions.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

A. Method/Procedure

The system follows a web-based approach where users interact with the application through a browser interface. The system processes uploaded DPR documents, extracts relevant information using NLP techniques, evaluates document quality, and predicts project risks through machine learning models.

3.1.1 System Design

The system is designed using a multi-layer client-server architecture. The user interacts with the frontend dashboard, while the backend server processes requests and communicates with machine learning models and the database.

3.1.2 Database Management

A PostgreSQL database is used to store DPR documents, extracted data, quality scores, risk predictions, and user feedback. The database is dynamically updated as new DPRs are uploaded and analyzed, enabling continuous learning and data management.

3.1.3 Document Upload and Processing

Users upload DPR documents through the web interface. The system extracts text from the documents using PDF processing and OCR techniques. The extracted data is cleaned and structured for further analysis.

3.1.4 NLP-Based Analysis

The system applies Natural Language Processing techniques to identify key sections such as project objectives, financial details, timelines, and technical information. This step converts unstructured document content into structured features.

3.1.5 Quality Evaluation

The system evaluates DPR quality based on criteria such as completeness, compliance, technical depth, and financial accuracy. A composite score is generated to represent the overall quality of the document.

3.1.6 Risk Prediction

Machine learning models such as XGBoost and LightGBM analyze extracted features to predict risks such as cost overruns, delays, and project feasibility issues.

3.1.7 Backend Development

The backend is developed using FastAPI, which manages API requests, processes DPR documents, performs NLP-based analysis, and integrates machine learning models for risk prediction. It ensures efficient communication between different system modules and the database.

3.1.8 Frontend Development

The frontend is developed using HTML, CSS, and JavaScript to provide a user-friendly dashboard for uploading DPR documents and viewing results. It interacts with the backend through REST APIs and dynamically displays outputs such as quality scores and risk predictions.

3.1.9 System Integration

All system components, including frontend,backend, database,NLP modules, and machine learning models, are integrated to provide a complete AI-powered DPR analysis system. This integration enables seamless data flow from document upload to final analysis and result visualization.

B. System Architecture

The AI-Powered DPR Analysis System follows a multi-layer architecture consisting of the client layer, application layer, core processing layer, and data layer. This architecture ensures scalability, maintainability, and efficient communication between system components. The client layer represents the user-facing interface, developed using HTML, CSS, and JavaScript. It allows users to upload DPR documents, view analysis results, and access dashboards.

The application layer is implemented using FastAPI, which handles API requests and manages communication between the frontend and backend services. It includes modules for document upload, NLP analysis, quality scoring, risk prediction, and learning.The core processing layer performs the main computations, including document parsing, NLP-based information extraction, quality scoring, compliance validation, and machine learning-based risk prediction using models like XGBoost and LightGBM. The data layer consists of a PostgreSQL database, which stores DPR documents, analysis results, quality scores, risk predictions, and training data. The database is continuously updated as users interact with the system, enabling incremental learning and improved predictions over time.

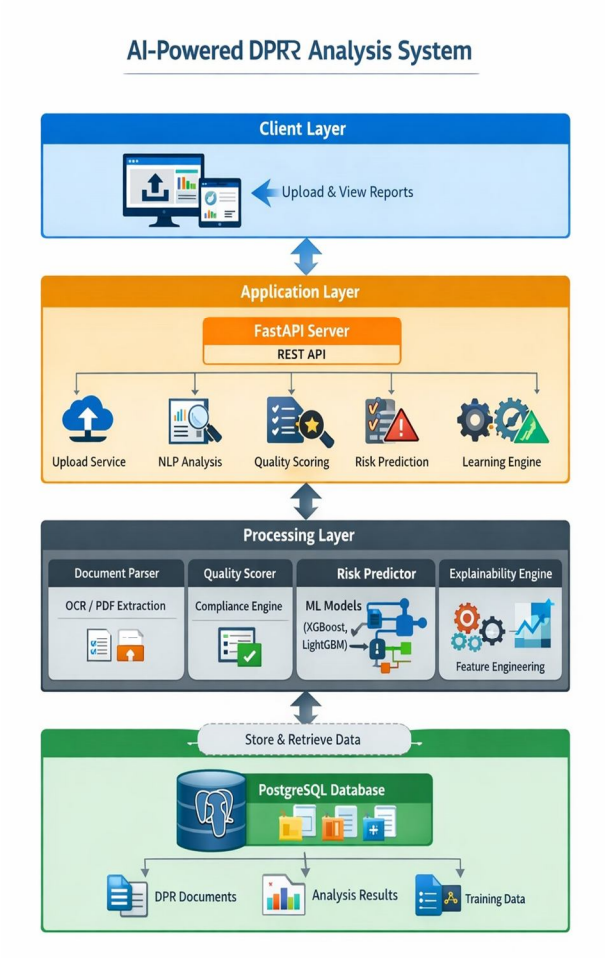


Fig3.2.1 System Architecture

C. Class Diagram

The class diagram represents the structure of the AI-Powered DPR Quality Assessment and Risk Prediction System by showing the main classes, their attributes, methods, and relationships. It explains how different components interact to perform DPR analysis, quality evaluation, risk prediction, and data management. The DPRDocument class manages document-related information. It contains attributes such as document_id, file_name, content, and upload_date. The methods include upload() and extractText() for handling document input.

The NLPProcessor class is responsible for analyzing the document using Natural Language Processing techniques. It includes methods such as extractSections(), extractEntities(), and processText() to convert unstructured data into structured information. The QualityScorer class evaluates the DPR based on various criteria. It contains attributes such as score and grade, and methods like calculateScore() to generate quality scores.

The ComplianceEngine class checks whether the DPR follows government rules and standards. It includes methods such as validateCompliance() to ensure adherence to required guidelines.

The RiskPredictor class handles risk analysis using machine learning models. It includes methods such as predictRisk() to estimate cost overrun, delay probability, and overall risk level. The ExplainabilityEngine class provides explanations for predictions. It generates human-readable insights using methods like generateExplanation().

The DatabaseManager class manages data storage. It includes methods such as saveData() and fetchData() for interacting with the PostgreSQL database. All these classes are interconnected. The DPRDocument is processed by the NLPProcessor, which provides data to the QualityScorer and RiskPredictor. The ComplianceEngine validates rules, and the ExplainabilityEngine generates insights. Finally, the DatabaseManager stores all results.

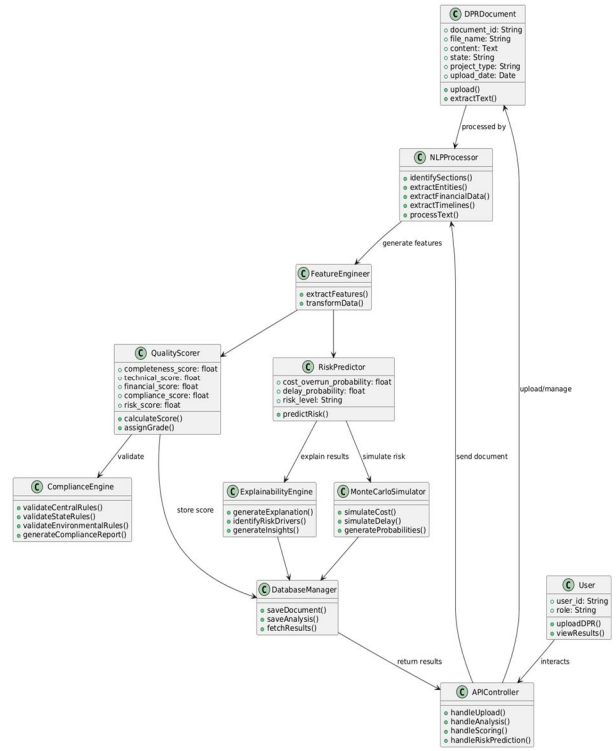


Fig3.3.1 Class Diagram

D. Sequence Diagram

The sequence diagram describes the interaction between different components of the DPR analysis system and shows how data flows from the user to the system and database during the analysis process. The process begins when the user uploads a DPR document through the web interface. The upload request is sent to the backend server, where the document is stored and processed.

The system extracts text from the document and passes it to the NLP module for analysis. The NLPProcessor identifies sections, extracts key information, and generates structured data. Next, the QualityScorer evaluates the document and generates a quality score, while the ComplianceEngine checks adherence to government standards. The processed data is then passed to the RiskPredictor, which uses machine learning models to predict risks such as cost overruns and delays.

The ExplainabilityEngine generates human-readable insights explaining the predictions. Finally, all results are stored in the database and displayed to the user through the interface. This sequence ensures smooth interaction between the user, system modules, and database, enabling efficient DPR analysis and decision support.

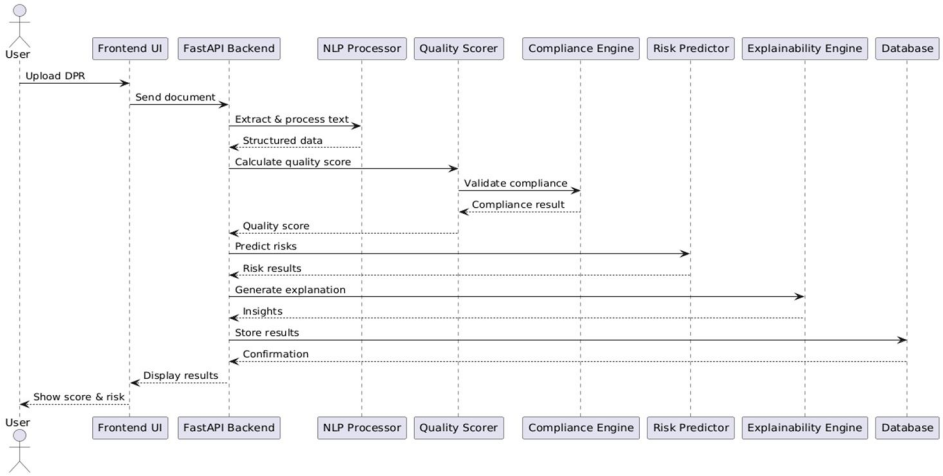


Fig 3.4.1 Sequence Diagram

IV. RESULTS AND FINDINGS

The AI-Powered DPR Quality Assessment and Risk Prediction System was successfully developed and tested. The system allows users to upload DPR documents and automatically analyze them through a web-based interface. The document processing module effectively extracts text and identifies key sections such as project objectives, financial details, and timelines. The NLP-based analysis accurately converts unstructured DPR content into structured data for further evaluation.

The quality scoring module successfully evaluates DPRs based on multiple criteria, generating a standardized score and grade. The risk prediction module accurately estimates probabilities of cost overruns and delays using machine learning models such as XGBoost and LightGBM. The system also provides explainable insights and probabilistic analysis, helping users understand the reasons behind risk predictions. Overall, the system reduces manual effort, improves evaluation accuracy, and provides a reliable decision-support tool for infrastructure project assessment.

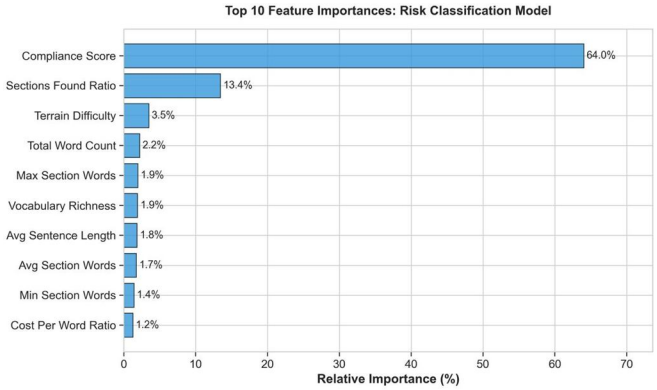


Fig 4.1 Top 10 Feature Importances for Risk Classification Model

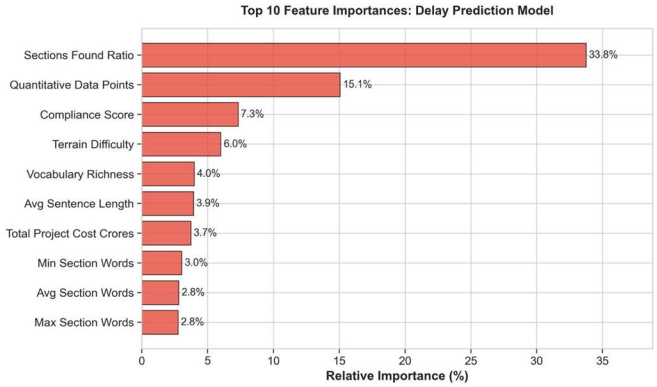


Figure 4.2: Top 10 Feature Importances for Delay Prediction Model

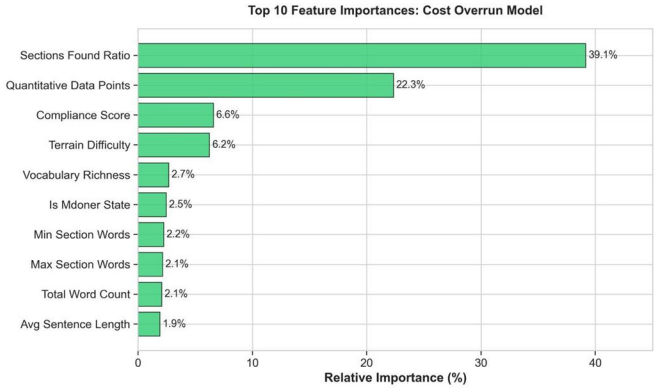


Figure 4.3: Performance Comparison of Prediction Models

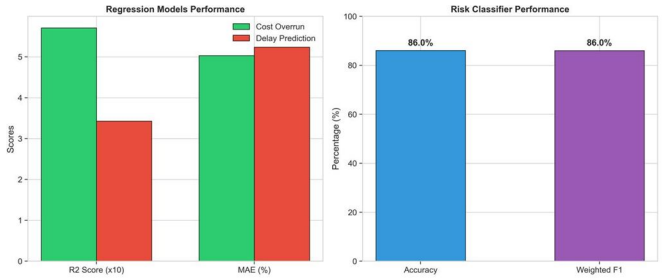


Figure 4.4: Top 10 Feature Importances for Cost Overrun Model

V. DISCUSSION

The results demonstrate that the integration of NLP and machine learning techniques significantly improves the efficiency of DPR evaluation compared to traditional manual methods. The system enables automated analysis of large and complex documents, reducing human effort and time required for review. The quality scoring mechanism ensures standardized evaluation, minimizing inconsistencies in decision-making. The risk prediction module provides early identification of potential issues, allowing decision-makers to take preventive actions.

The inclusion of Explainable AI improves transparency by providing clear reasoning behind predictions. However, the system's performance depends on the availability and quality of training data. Additionally, variations in DPR formats may affect extraction accuracy. Future improvements such as better training datasets, advanced NLP models, and integration with government systems can further enhance system performance and reliability.

VI. CONCLUSION AND FUTURE SCOPE

The proposed AI-Powered DPR Analysis System provides an efficient and intelligent solution for evaluating infrastructure project reports. By automating document analysis, quality assessment, compliance validation, and risk prediction, the system improves accuracy, reduces manual effort, and supports data-driven decision-making.

The system helps in identifying potential risks early, ensuring better project planning and reducing cost overruns and delays. Overall, it enhances transparency, consistency, and efficiency in government project evaluation.

In the future, the system can be enhanced by integrating advanced explainability techniques such as SHAP, supporting multi-language DPR analysis, and deploying the system on cloud platforms for large-scale usage. Integration with government databases and real-time project monitoring systems can further improve its effectiveness and applicability.

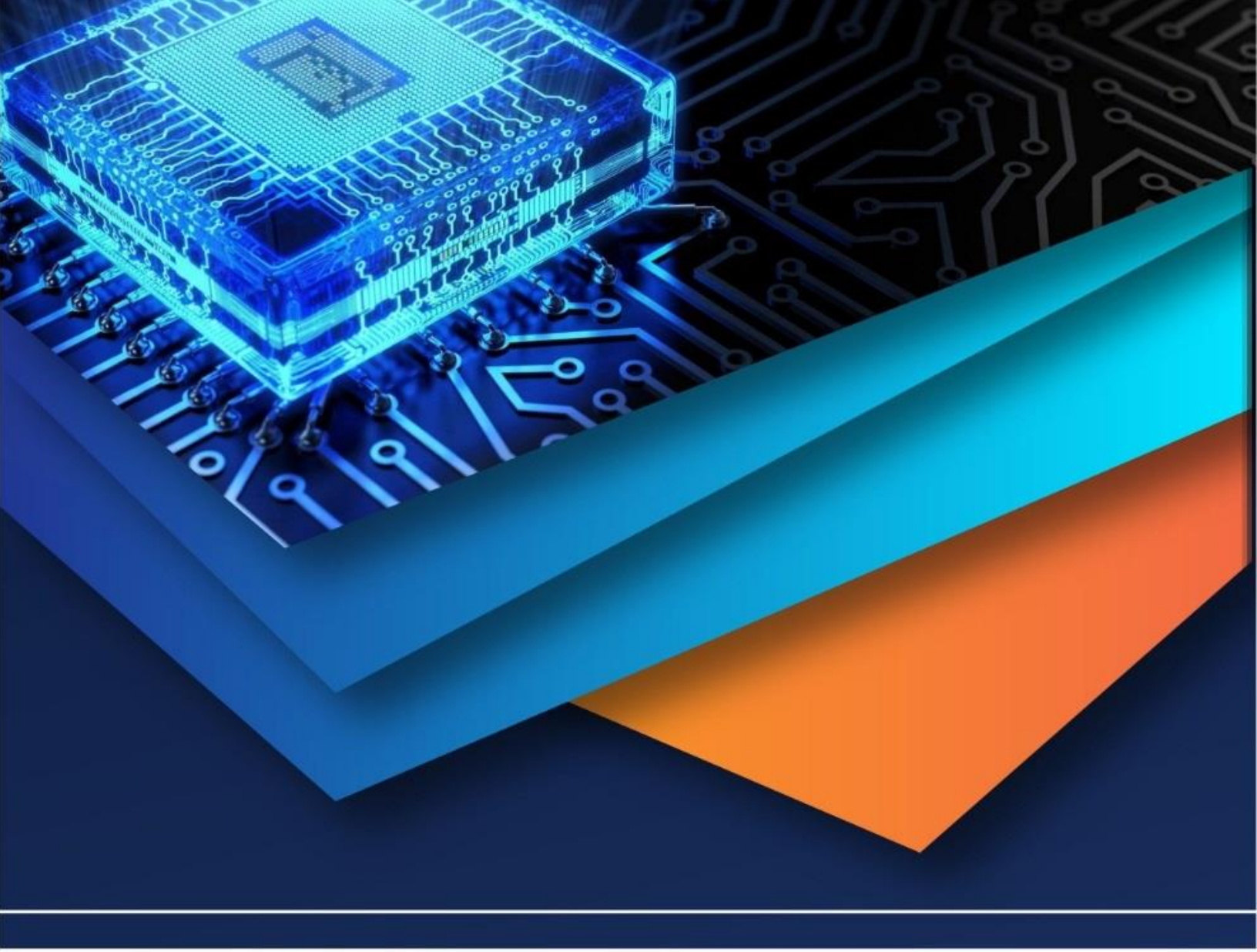
VII. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the faculty members and project guide for their valuable guidance and continuous support throughout the development of this project. Their suggestions, insights, and encouragement played a crucial role in the successful completion of the AI-Powered DPR Quality Assessment and Risk Prediction System.

We also extend our thanks to our institution for providing the necessary resources, tools, and infrastructure required to carry out this research work effectively.

REFERENCES

[1] AI-Based Project Risk Prediction using Machine Learning, IEEE (2023)
[2] Natural Language Processing for Document Analysis, Elsevier (2022)
[3] Intelligent Document Processing using OCR and NLP, IEEE Access (2023)
[4] XGBoost and LightGBM for Predictive Analytics in Infrastructure Projects (2024)
[5] Explainable AI for Decision Support Systems, IEEE (2023)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY